(72) Inventors; and
(75) Inventors/Applicants (for US only): ANDERSON, Malcolm [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). KEEN, Nicholas, John [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). PANNIFER, Andrew, David, Bruce [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). PAUPTIT, Richard, Alexander [NL/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). ROWSELL, Sian [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB).

(74) Agent: ASTRAZENECA; Global Intellectual Property, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG (GB).

(54) Title: CRYSTAL STRUCTURE OF ENZYME AND USES THEREOF

(57) Abstract: The invention provides crystalline forms of a polypeptide corresponding to the catalytic domain of Aurora kinase. The active site ATP binding pocket is defined by its amino acid residues and their atomic coordinates. This structure may be used to select or design chemical modulators of Aurora kinase, particularly Aurora inhibitors. These modulators may be used to treat diseases of cell proliferation, e.g. cancer.

-1-
CRYSTAL STRUCTURE OF ENZYME AND USES THEREOF

Field of the Invention

This invention relates to crystallised Aurora A kinase and the use of its three-
5 dimensional structure to investigate Aurora kinase homologues and to design Aurora kinase

modulators.

Background of the Invention

Proteins such as enzymes involved in physiological and pathological processes are

10 important targets in the development of pharmaceutical compounds and treatments.

Knowledge of the three dimensional (tertiary) structure of proteins allows the rational design

of mimics or modulators of such proteins. By searching structural databases using structural

parameters derived from the protein of interest, it is possible to select molecular structures that

may mimic or interact with these parameters. It is then possible to synthesise the selected

15 molecular structure and test its activity. Alternatively, the structural parameters derived from

the protein of interest may be used to design and synthesise a mimic or modulator with the

desired activity. Such mimics or modulators may be useful as therapeutic agents for treating

certain diseases. For example, WO98/07835 discloses crystal structures of a protein tyrosine

kinase optionally complexed with one or more compounds. The atomic coordinates of the

20 enzyme structures and any of the bound compounds are used to determine the three-

dimensional structures of kinases with unknown structure and to identify modulators of kinase

functions. As another example, WO99/01476 discloses the crystal structures of anti-Factor IX

Fab fragments (antibodies) and their use to identify and design new anticoagulant agents.

Knowledge of the three-dimensional structure of a protein is essential for the rational

25 design of mimics or modulators of that protein. Lack of structural knowledge is a barrier to

the development of new mimics or modulators that may have extremely useful pharmaceutical

properties.

In Eukaryotes, the cell cycle is largely controlled by an ordered cascade of protein

phosphorylation. Several families of protein kinases that play critical roles in this cascade

30 have now been identified. The activity of many of these kinases is increased in human

tumours when compared to normal tissue. This can occur by either increased levels of

expression of the protein (as a result of gene amplification for example), or by changes in

expression of co-activators or inhibitory proteins.

-2-

The first identified, and most widely studied of these cell cycle regulators have been the cyclin-dependent kinases (or CDKs). Activity of specific CDKs at specific times is essential for both initiation and coordinated progress through the cell cycle. For example, the CDK4 protein appears to control entry into the cell cycle (the G0-G1-S transition) by

5   phosphorylating the retinoblastoma gene product pRb. This stimulates the release of the transcription factor E2F from pRb, which then acts to increase the transcription of genes necessary for entry into S phase. The catalytic activity of CDK4 is stimulated by binding to a partner protein, Cyclin D1. One of the first demonstrations of a direct link between cancer and the cell cycle was made with the observation that the Cyclin D1 gene was amplified and cyclin

10 D1 protein levels increased (and hence the activity of CDK4 increased) in many human tumours (Reviewed in Sherr, 1996, Science 274: 1672-1677; Pines, 1995, Seminars in Cancer Biology 6: 63-72). Other studies have shown that negative regulators of CDK function are frequently down-regulated or deleted in human tumours, again leading to inappropriate activation of these kinases (Loda et al., 1997, Nature Medicine 3(2): 231-234; Gemma et al.,

15 1996, International Journal of Cancer 68(5): 605-11; Elledge et al. 1996, Trends in Cell Biology 6; 388-392).

More recently, protein kinases that are structurally distinct from the CDK family have been identified which play critical roles in regulating the cell cycle and which also appear to be important in oncogenesis. These include the newly-identified human homologues of the

20 *Drosophila* Aurora and *S. cerevisiae* Ipl1 proteins. *Drosophila* Aurora and *S. cerevisiae* Ipl1, which are highly homologous at the amino acid sequence level, encode serine/threonine protein kinases. Both Aurora and Ipl1 are known to be involved in controlling the transition from the G2 phase of the cell cycle through mitosis, centrosome function, formation of a mitotic spindle and proper chromosome separation / segregation into daughter cells. The three

25 human homologues of these genes, termed Aurora A, B and C, encode cell cycle regulated protein kinases. These show a peak of expression and kinase activity at the G2/M boundary (Aurora A, C) and in mitosis and cytokinesis (Aurora B). Several observations implicate the involvement of human Aurora proteins, in particular Aurora A in cancer. The Aurora A gene maps to chromosome 20q13, a region that is frequently amplified in human tumours including

30 both breast and colon tumours. Aurora A may be the major target gene of this amplicon, since Aurora A DNA is amplified and Aurora A mRNA over expressed in greater than 50% of primary human colorectal cancers. In these tumours Aurora A protein levels appear greatly elevated compared to adjacent normal tissue. In addition, transfection of rodent fibroblasts

-3-

with human Aurora A leads to transformation, conferring the ability to grow in soft agar and

form tumours in nude mice (Bischoff et al., 1998, The EMBO Journal. 117(11): 3052-3065).

Other work has shown that artificial over expression of Aurora A leads to an increase in

centrosome number and an increase in aneuploidy (Zhou et al., 1998, Nature Genetics. 20(2):

5   189-93).

Importantly, it has also been demonstrated that abrogation of Aurora A expression and

function by antisense oligonucleotide treatment of human tumour cell lines (Bischoff and

Ploughman, 1999, Trends in Cell Biology, 9(11): 454-459 or by a small molecule inhibitor of

Aurora A kinase activity (Keen et al. 2001, poster #2455, American Association for Cancer

10 Research annual meeting, New Orleans USA) leads to defects in mitosis, cell cycle arrest and

exerts an antiproliferative effect in these tumour cell lines. This indicates that inhibition of

the function of Aurora A will have an antiproliferative effect that may be useful in the

treatment of human tumours and other hyperproliferative diseases.

In order to design inhibitors of Aurora A kinase, it is necessary to know the three-

15 dimensional structure of Aurora A kinase, in complex with various lead compounds. To date,

the three-dimensional structure of Aurora A kinase has not been available. Further, it has not

been possible to obtain crystals of any part of Aurora of sufficient quality to allow

determination of the structure of the kinase domain including the site of inhibition.

20 **Summary of the Invention**

The present invention relates to the previously unknown three-dimensional structure of

human Aurora A kinase. As described herein, the Applicants have overcome the difficulties

encountered by others and have produced crystals of the Aurora A kinase catalytic domain

that are of sufficient quality to determine the three-dimensional structure of the protein by X-

25 ray diffraction methods. In addition, the Applicants have determined the three-dimensional

crystal structure of the kinase catalytic domain of Aurora A kinase in a complex with the ATP

analogue AMP-PNP, as well as the three-dimensional crystal structure of the Aurora A kinase

catalytic domain in complex with a synthetic inhibitor. There is a clear need for this structural

information to enable identification and structure-based design of new Aurora kinase

30 modulators (particularly inhibitors) for the treatment of various diseases or conditions and in

particular diseases of cell proliferation such as cancer. The methods described herein allow the

determination of the three-dimensional structures of Aurora A kinase, as well as other Aurora

-4-

kinases, in complex with numerous inhibitors of interest to aid in the rational design of modulators that will treat diseases of cell proliferation.

Brief Description of the Drawings

Figure 1 is a schematic representation of the structure of the [T287D] Aurora A complex with AMP-PNP. The inhibitor has 2 conformations.

Figure 2a is a schematic representation of the structure of Aurora A in complex with a synthetic inhibitor drawn in approximately the same orientation as Figure 1.

Figure 2b is a schematic representation of Aurora A in complex with a synthetic inhibitor, rotated so as to show the extended inhibitor occupying a long active site binding pocket.

Figure 3 is a graph of the activity of [T287D]Aurora A as a function of pH.

Detailed Description of the Invention

This invention relates to crystals of Aurora A kinase and the use of the three-dimensional structure to investigate Aurora kinase homologues and to design Aurora kinase modulators (preferably inhibitors). It further relates to crystals of Aurora kinase, particularly Aurora A kinase, or the catalytic portion thereof, complexed or uncomplexed as described, of sufficient quality to determine the three dimensional (tertiary) structure of the polypeptide by X-ray diffraction methods.

According to a first aspect of the invention, the Applicants provide two crystalline forms of a polypeptide comprising the catalytic domain of Aurora A kinase. One crystalline form is obtained when we crystallise [T287D]Aurora A(122-396) in the presence of the ATP-analogue AMP-PNP. The second crystalline form is obtained when we crystallise GSHM-[T287D]Aurora A(122-400) in the presence of a synthetic inhibitor. (Amino acid residues in Aurora A are numbered by taking the first amino acid immediately after the initial methionine as amino acid number one). In one embodiment, the first crystalline form has the space group $P3_221$. In another embodiment, the first crystalline form has the unit cell dimensions a = b = 86.55, c = 78.34 Å, $\alpha = \beta = 90$ and $\gamma = 120°$. In another embodiment, the second crystalline form has space group $P2_1$. In another embodiment, the second crystalline form has the unit

-5-

cell dimensions a = 52.6, b = 88.4, c = 67.8 Å, α= γ = 90 and β = 90.01°. In another

embodiment, these crystalline forms are described by three-dimensional sets of x,y,z-

coordinates (Tables 1 and 2) for each atom in the complex representing the unique repeating

motif in the crystal. Table 1 contains the coordinates for the complex molecule in the first

5   crystalline form; Table 2 contains the coordinates for two independent complex molecules in

the asymmetric unit (smallest unique repeating unit) in the second crystalline form. In another

embodiment, these crystalline forms contain a numerical definition of a binding site,

approximated by the set of all residues within a 5 Å contact distance from any atom in either

inhibitor. The binding site is defined by the x,y,z-coordinates of atoms in the set of amino acid

10   residues (set A) given by the list Arg136, Leu138, Gly139, Lys140, Gly141, Val146, Ala 159,

Lys161, Leu163, Val177, Glu180, Val181, Ile183, Gln184, Leu193, Leu195, Leu207,

Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259,

Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278, and His279,

the atomic coordinates being listed in Tables 1 and 2. The binding site is may be defined in

15   any alternate crystalline form, homologue, variant or mutant wherein the binding site has a

root mean square deviation from all atoms of the amino acid residues of not more than 1.0 Å

from a least-flexible subset (set B) of the binding site that includes the amino acid residues

Arg136, Leu138, Gly139, Val146, Ala159, Lys161, Leu163, Ile183, Gln184, Leu193, Leu195,

Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219,

20   Glu259, Asn260 and Leu262, each having coordinates as described in Tables 1 and 2.

In another embodiment, the first crystalline form comprises a binding site defined by

amino acid residues Leu138, Gly139, Val146, Lys161, Val177, Arg178, Arg179, Glu180,

Val181, Glu182, Ile183, Gln184, Leu193, Leu209, Tyr211, Ala212, Gly215, Thr216, Glu259,

Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278 and His 279,

25   each having the coordinated listed in Table 1a. An alternative crystalline form, homologue,

variant or mutant wherein the binding site has a root mean square deviation from the

backbone atoms of the amino acid residues of not more than 1.5 Å, and preferably not more

than 1.0 Å is also provided.

In another embodiment, the crystalline forms additionally comprise Aurora kinase

30   inhibitors in complex with the catalytic domain of Aurora kinase including any of the above

embodiments of the crystalline form.

-6-

Another aspect of the invention relates to a method of designing an Aurora chemical modulator using the atomic coordinates of a crystalline form according to any of the above embodiments.

Another aspect of the invention relates to a method of selecting an Aurora chemical
5 modulator using the atomic coordinates of a crystalline form according to any of the above embodiments.

Another aspect of the invention relates to a method of designing or selecting an Aurora chemical modulator using the atomic coordinates of any other protein, e.g. PKA, which has been shown by this invention to have structural similarity to Aurora.

10 Another aspect of the invention relates to a method of designing an Aurora protein using the atomic coordinates of a crystalline form according to any of the above embodiments.

Another aspect of the invention relates to a method of designing or selecting an Aurora modulator comprising the steps of:

exploring the atomic coordinates of Aurora (Tables 1 and 2) for information on the
15 three-dimensional characteristics of the protein surface;

arriving at an alternative overlapping or non-overlapping binding pocket to the active site ATP binding pocket; and

selecting or designing an Aurora modulator using the binding pocket information.

Another aspect of the invention relates to a method of determining the three-
20 dimensional structure of a crystal form of Aurora kinase, referred to as a second or new crystal or crystal form of Aurora kinase, comprising the step of applying difference Fourier or molecular replacement methods using the atomic coordinates of an original crystal of Aurora kinase (from Table 1 or 2) to model the structure of a new crystal, wherein the active site ATP binding pocket of the new crystal is equivalent to that in the first crystal. In a specific
25 embodiment, the invention is a method of determining the three-dimensional structure of a crystal form of Aurora kinase A comprising the step of applying difference Fourier or molecular replacement methods using the atomic coordinates of an original (first) crystal of Aurora kinases (from Table 1 or 2) to model the structure of a new crystal or new crystal form of Aurora kinase A, wherein the active site ATP binding pocket of the new crystal is
30 equivalent to that in the original (first) crystal.

In particular provided herein are crystalline forms of a polypeptide including the catalytic domain of an Aurora A protein. The catalytic domain may be found within the complete protein or within a fragment of the protein. The catalytic domain may be also

-7-

derived from a wild-type Aurora A enzyme or from an Aurora A mutant, homologue or variant. A mutant is a wild type Aurora A protein having one or more changes in its amino acid sequence. An Aurora mutant may have the same activity as the wild type protein, may have modified activity or may be inactive. A variant is a wild type or mutant protein having

5 one or more portions of its sequence removed, or an additional sequence or sequences added, so that the variant is a different length from the wild type or mutant protein. A variant usually has the same activity as the original wild type or mutant protein. A homologue is a related protein in which some parts of the amino acid sequence are the same as in the original protein. Aurora B and Aurora C, for example, are homologues of Aurora A.

10   The invention relates to crystals of sufficient quality to determine the three dimensional structure to high resolution of any portion, mutant, variant or homologue of Aurora A involving the catalytic domain.

      According to a further aspect of the invention, we provide crystalline forms of a polypeptide containing the Aurora A catalytic domain in complex with small molecular

15 weight inhibitor molecules. For example, the inhibitor molecule might be a non-hydrolysable analogue of ATP. Such analogues include, for example, formula I (AMP-PNP). As another example, the inhibitor might be a molecule synthesised chemically. Such molecules include, for example, formula II.

Formula I: AMP-PNP

20
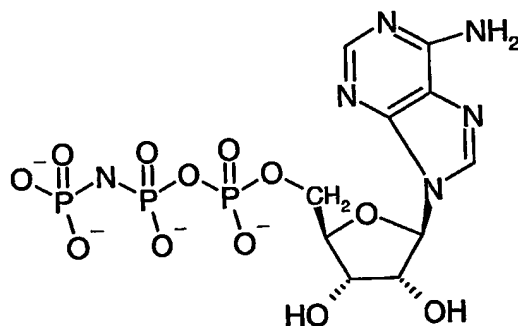
-8-
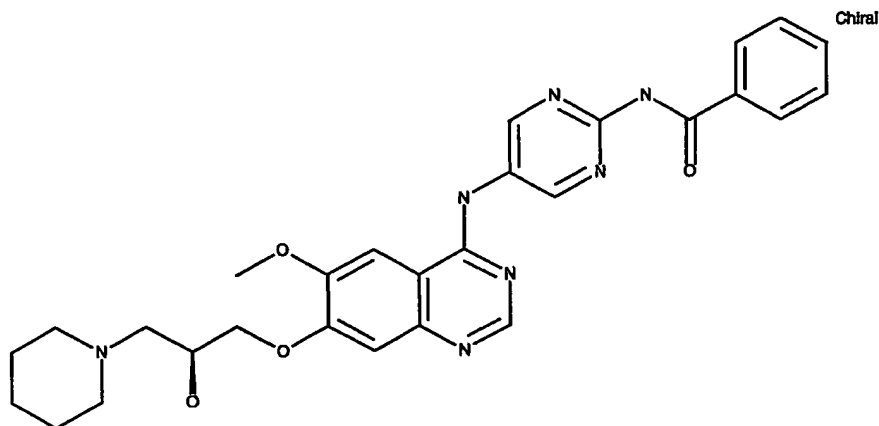
Formula II

Another aspect of the invention is the unique shape of the active site ATP binding

pocket in Aurora. Using X-ray crystallography, we have determined the three-dimensional

5 molecular structure of an Aurora A catalytic domain. Resulting from this, we have

determined the unique shape of an Aurora A active site ATP binding pocket (defined by the

atomic coordinates of its constituent amino acids). Furthermore, we have determined the

spatial arrangement of an Aurora A substrate analogue and an inhibitor molecule relative to

the Aurora A active site binding pocket. This structural information can be stored on a

10 computer-readable medium and may be used for rational drug design.

One of the difficulties in studying kinases in general is obtaining active protein. In

order to be activated, certain kinases need to be phosphorylated at one or more key amino acid

residues. It may be experimentally difficult to obtain 100% pure phosphorylated protein.

Different phosphorylation states may have different conformations. Those in the art realise

15 that such heterogeneities in the protein sample can severely impede the ability to form large

well-ordered crystals. In Aurora A, phosphorylation of Thr 287 is necessary for activation of

the kinase. Replacement of Thr-287 by Asp (an Aurora A mutant called [T287D] Aurora A)

provides a mimic of the active protein which can be provided as a homogeneous sample. The

[T287D]Aurora A mutant is constitutively active. Thus, preparation of this mutant

20 conveniently addresses both issues of activity and crystallisability.

One of the major hurdles in the crystallisation of multidomain proteins is their

flexibility. To increase the chances of crystallising Aurora A, an enzyme construct limited to

the catalytic domain was used. This provided a more rigid and compact domain. Catalytic

domain constructs can be designed by comparing the amino acid sequence to other kinases of

25 known structure, and defining start and end residues for the polypeptide encompassing the

-9-

Aurora A catalytic domain by analogy. This gives numerous possible construct variants, which include the catalytic domain. In order to increase further the chances of crystallising Aurora A, experimental evidence was sought as to which catalytic domain construct would be the most compact while retaining integrity as a folding unit. Limited proteolysis was carried

5 out using endoproteinase Glu-C from Staphylococcus aureus V8 on the catalytic domain. This defined the catalytic domain boundaries to be within residues 122 to 396. Other similar constructs may be obtained through similar procedures, using, for example, different proteases for the limited proteolysis experiment. Such a procedure is exemplified by our preparation, crystallisation and determination of the structure of two Aurora A catalytic domain

10 polypeptides. The structure of [T287D]Aurora A(122-396) in complex with the non-hydrolysable ATP analogue AMP-PNP is shown in Fig. 1. The structure of GSHM-[T287D]Aurora A(122-400) in complex with the synthetic Aurora inhibitor of formula II is shown in Fig. 2a and 2b.

The AMP-PNP molecule occupies a cleft between the N-terminal domain (residues

15 125 to 208) and the C-terminal domain (residues 215 to 374). Comparison with other kinases demonstrates that this cleft represents a portion of the ATP binding site. Therefore, we have identified the active site ATP binding pocket of Aurora. The electron density shows evidence for the AMP-PNP adopting a dual conformation. In both conformations, the adenine ring and ribose moiety occupy similar pockets with the adenine nitrogen atoms N1 and N6 making

20 classical interactions with main chain atoms in the hinge region (residues 209 to 214) of the enzyme. N1 forms a hydrogen bond with the main chain nitrogen of Ala-212 while N6 forms a hydrogen bond to the peptide carbonyl group of Glu-210. However, torsion angle differences elsewhere in the molecule allow the alpha and beta phosphate groups to occupy alternative pockets. No electron density is apparent in either conformation for the gamma

25 phosphate group of the AMP-PNP molecule. In conformation 1, the beta phosphate group makes polar interactions with the O oxygen atoms of Ser 277 and the side-chain of Asn260, while in conformation 2, the beta phosphate makes polar interactions with the amide carbonyl of –Glu-259 and with a water molecule (Wat-542 in this structure).

From the three-dimensional structure that we have determined for [T287D] Aurora A,

30 we establish that the AMP-PNP binding pocket, which is the active site ATP binding pocket, is uniquely defined by the atomic co-ordinates of its constituent amino acid residues, the coordinates being listed in Tables 1 and 2. An equivalent ATP binding pocket may also be defined having the same co-ordinates as detailed in Table 1 and with the same constitutent

-10-

amino acid residues except that Lys140 and Gly141 in Table 1 and replaced with Ala140 and Ala141, whereby such a table is referred to hereonin as Table 1a.

5

10

15

20

25

30

-104-

The shape of the ATP binding pocket is defined by the atomic coordinates of the atoms in the amino-acid residues in Tables 1 and 2. Table 1 lists the atomic coordinates for [T287D] Aurora A(122-396) catalytic domain, together with the AMP-PNP molecule, in Protein Data Bank (PDB) format, as determined from the first crystalline form. Table 2-lists

5 the atomic coordinates for the two independent molecules of the GSHM-[T287D]Aurora A (122-400) catalytic domain, together with the inhibitor of formula II, in PDB format, as determined from the second crystalline form. The atomic coordinates are listed in those lines that begin with the code ATOM or HETATM, one atom per line. Following the code are: the unique atom number; the atom name; the amino acid residue name; the protein chain

10 identifier; the amino acid residue number; the atomic coordinates x, y, and z in orthogonal Angstrom space; the atomic occupancy factor; the atomic temperature factor; the chain identifier; and the atom type. The atomic coordinates of the ATP analogue AMP-PNP carry the residue name of-ANP. Solvent water molecules carry the residue name of HOH, and a citrate and a bound phosphate derived from the crystallisation buffer carry the residue name of

15 FRA. In the inhibitor complex the inhibitor molecules carry the residue name of FRA.

It is possible to reproduce the shape of the [T287D]Aurora A active site binding pocket through carrying out similar structure determinations with minor variations in the experimental conditions (including variations in construct such as mutants, variants and homologues, variations in crystallisation conditions, crystal form, trial model used in

20 molecular replacement, etc.). Different experiments may give rise to apparently different co-ordinates, but those in the art will realise that two apparently different sets of coordinates for the same or similar proteins can be shown to be equivalent by superposition of the molecules. For example, the coordinates in Tables 1 and 2 are different numerically. But following superposition they can be seen to describe the same molecule. It will be appreciated that,

25 according to accepted practice, the atomic coordinates may vary within certain limits due to experimental variation. Such variation includes standard experimental error (coordinates determined for the same construct may vary somewhat, for example within 0.3 Å) and other variation (for example, coordinates of Aurora mutants, variants, or homologues). The co-ordinates of the active site ATP binding site may also differ upon introduction of a different

30 small molecule inhibitor, where flexible portions of the binding site adopt a new conformation specific to a type of inhibitor. For example, following superposition, the protein coordinates in Table 1 are seen to be marginally different to those in Table 2, as a result of flexible portions of the protein being influenced by the presence of a different inhibitor. This constitutes a

-105-

modification of the active site ATP binding site rather than the creation of a new site. Those in the art will realise that kinases in general have flexible active sites, and adopt a number of biologically relevant conformations related to the state of catalytic activation. Therefore, for the purposes of differentiating the shape of the active site ATP-binding pocket from that in

5 other kinases, the binding pocket is best defined by a subset of amino acids that are least affected by flexible protein responses to inhibitor binding. Thus, a protein can be said to have the Aurora active site described here if, following superposition, the positions of all atoms in the active site residues in set B, i.e. Arg136, Leu138, Gly139, Val146, Ala159, Lys161, Leu163, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213,

10 Leu214, Gly215, Thr216, Arg219, Glu259, Asn260 and Leu262 or their equivalents, are within a root mean square deviation of 1.0 Å of the coordinates of these amino acid residues given in Tables 1 and 2. An equivalent residue is an amino acid residue in any Aurora mutant, variant, or homologue that occurs at one of the amino acid sequence positions in Tables 1 or 2 – if the residue is not identical, only the N, Cα, Cβ, C, O atoms may be sensibly included in

15 the rmsd calculation. It is also understood that if equivalent residues are not present in a particular variant or homologue, then they are omitted from the calculation of the average distance.

The criterion of 1.0 Å is intended to be large enough to allow the types of variations described above, yet small enough to discriminate between the active sites of Aurora kinases

20 and other kinases. That this criterion is reasonable is illustrated in Table 3, which compares [T287D]Aurora A to one of the most closely related kinases, PKA.

			1atp	1l3r	1bx6	1apm	1ydt	1cmk
1- 2	0.85	1-PKA	1.32	1.32	1.29	1.24	1.28	1.17
1 - 3	0.67	2-PKA	1.19	1.14	1.08	1.13	1.13	1.13
2 - 3	0.42	3-PKA	1.14	1.08	1.09	1.08	1.10	1.21

Table 3: rms deviations in Å between all atoms of set B amino acids in the active site.

25 The top row refers to the PDB codes for 6 entries of a different kinase, protein kinase A (PKA). The bold numbers refer to the Tables 1 and 2 which contain Aurora coordinates. Thus, when independent structure determinations of Aurora are compared (1-2, 1-3, 2-3) the rmsd is

-106-

less than 1 Å, whereas when Aurora is compared to PKA (6 independent structures) the rmsd
is greater than 1 Å.

Thus, according to a further aspect of the invention, we provide the shape of the active

5   site ATP binding pocket in Aurora protein kinase as defined by the atomic coordinates given

in Tables 1 and 2 or by equivalent coordinates. Equivalent coordinates are those for which the

subset of least flexible residues (set B) have atomic positions on average within 1.0 Å of those

in the Aurora active site ATP binding pocket as defined by the coordinates in Tables 1 and 2.

According to a further aspect of the invention we provide a method to determine or

10   design the three-dimensional structure of a crystal form of Aurora (including Aurora A

homologues, variants, mutants, and inhibitor complexes) by using a particular Aurora A

catalytic domain structure. The atomic co-ordinates of an Aurora A crystal may be used to

model the structure of a second Aurora crystal by difference Fourier or molecular replacement

methods.

15   The crystal structure of the Aurora A kinase catalytic domain described herein can be

used to model the three-dimensional structures of other Aurora kinases. Furthermore,

alternative methods of determining three-dimensional structure that do not rely on X-ray

diffraction techniques and thus do not require crystallization of the protein, such as NMR

techniques, are simplified if a model of the structure is available for refinement using the

20   additional data gathered by the alternative technique. Thus, definition of the three-

dimensional structure of the catalytic domain of Aurora A kinase enables one of skill in the art

to determine the structure of the catalytic domains of other Aurora kinases.

Knowledge of the three-dimensional structure of the catalytic domain of Aurora A

kinase provides a means for investigating the mechanism of action of the protein and tools for

25   identifying inhibitors of its function. Knowledge of the three-dimensional structure of the

catalytic domain of Aurora A kinase allows one to design molecules capable of binding

thereto, including molecules which are capable of inhibiting (partially or completely) the

activity of Aurora A kinase.

Illustrative crystalline forms of polypeptides of this invention having various

30   physicochemical characteristics are disclosed herein. Preferred crystalline forms invention are

capable of diffracting x-rays to a resolution of better than about 3.5 Å, and more preferably to

a resolution of 3.0 Å or better, and even more preferably to a resolution of 2.2 Å or better, and

are useful for determining the three-dimensional structure of the material.

-107-

Crystalline compositions of this invention specifically include those in which the crystals comprise Aurora kinase family proteins characterized by the structural coordinates set forth in any of the accompanying tables or characterized by coordinates having a root mean square deviation therefrom, with respect to backbone atoms of amino acids given in the

5   Tables, of 1.5 Å or less. Crystalline compositions of this invention also include those in which the crystals comprise Aurora kinase family proteins characterized by having a binding site defined by the x,y,z-coordinates of atoms in the set of amino acid residues (set A) given by the list Arg136, Leu138, Gly139, Lys140, Gly141, Val146, Lys161, Leu163, Val177, Glu180, Val181, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212,

10   Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278, and His279, the atomic coordinates being listed in Tables 1 and 2. Further, crystalline forms of polypeptides of this invention also include those in which the crystals comprise Aurora kinase family proteins in which the binding site is defined by the x,y,z-coordinates of atoms in the set of amino acid residues (set B) given by the

15   list Arg136, Leu138, Gly139, Val146, Ala159, Lys161, Leu163, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260 and Leu262 or their equivalent, are within a root mean square deviation of 1.0 Å of the coordinates of these amino acid residues given in Tables 1 and 2.

Structural coordinates of a crystalline composition of this invention may be stored in a

20   machine-readable form on a machine-readable storage medium, such as a computer hard drive, diskette, DAT tape, for display as a three-dimensional shape or for other uses involving computer-assisted manipulation of, or computation based on, the structural coordinates or the three-dimensional structures they define. For example, data defining the three dimensional structure of a protein of the Aurora kinase family, or portions or structurally similar

25   homologues of such proteins, may be stored in a machine-readable storage medium and displayed as a three-dimensional representation of the protein structure, typically using a computer capable of reading the data from said storage medium and programmed with instructions for creating the representation from such data. This invention thus encompasses a machine, such as a computer, having a memory which contains data representing the structural

30   coordinates of a crystalline composition of this invention, such as the coordinates set forth in Tables 1 and 2, together with additional optional data and instructions for manipulating such data. Such data may be used for a variety of purposes, such as the elucidation of other related structures and drug discovery.

-108-

For example, a first set of such machine readable data may be combined with a second set of machine-readable data using a machine programmed with instructions for using the first data set and the second data set to determine at least a portion of the coordinates corresponding to the second set of machine-readable data. For instance, the first set of data may comprise a Fourier transform of at least a portion of the coordinates for Aurora kinase proteins set forth in Tables 1 and 2, while the second data set may comprise X-ray diffraction data of a molecule or molecular complex.

More specifically, one of the objects of this invention is to provide three-dimensional structural information on new complexes of Aurora kinase family members (e.g., complexed with an ATP analogue or an inhibitor, such as a synthetic inhibitor), new Aurora kinase family members and variants of any of the foregoing. The structural coordinates of a crystalline composition of this invention, or portions thereof, can be used to solve, e.g. by molecular replacement, the three dimensional structure of a crystalline form of such a polypeptide or polypeptide complex. Doing so involves obtaining x-ray diffraction data for crystals of the polypeptide or polypeptide complex (e.g., in complex with an ATP analogue or an inhibitor, such as a synthetic inhibitor) for which one wishes to determine the three dimensional structure. The three-dimensional structure of that polypeptide or complex is determined by analyzing the x-ray diffraction data using molecular replacement techniques with reference to the structural coordinates provided. For example, molecular replacement can use a molecule having a known structure as a starting point to model the structure of an unknown crystalline sample. This technique is based on the principle that two molecules which have similar structures, orientations and positions in the unit cell diffract similarly. The term "molecular replacement" refers to a method that involves generating a preliminary model of a crystal whose atomic coordinates are not known, by orienting and positioning a related molecule whose atomic coordinates are known. Phases are then calculated from this model and combined with observed amplitudes to give an approximate Fourier synthesis of the structure whose coordinates are unknown. Molecular replacement involves positioning the known structure in the unit cell in the same location and orientation as the unknown structure. Once positioned, the atoms of the known structure in the unit cell are used to calculate the structure factors that would result from a hypothetical diffraction experiment. This involves rotating the known structure in the six dimensions (three angular and three spatial dimensions) until alignment of the known structure with the experimental data is achieved. This approximate structure can be refined to yield a more accurate and often higher resolution structure using

-109-

various refinement techniques. For instance, the resultant model for the structure defined by

the experimental data may be subjected to rigid body refinement in which the model is

subjected to limited additional rotation in the six dimensions yielding positioning shifts of

under about 5%. The refined model may then be further refined using other known

5 refinement methods.

For example, one may use molecular replacement to exploit a set of coordinates such

as set forth in Table 1 or Table 2 to determine the structure of the catalytic domain of Aurora

kinase in complex with other than ATP-PNP or the inhibitor of formula II.

The present invention also relates to designing and, optionally producing, a homologue

10 of Aurora kinase, such as a homologue of Aurora kinase A, that mimics the three-dimensional

structure of the Aurora kinase. The method comprises:

(i)   determining the three-dimensional coordinates of atoms of an Aurora kinase;

(ii)   providing a computer having a memory means, a data input means, a visual

display means, said memory means containing three-dimensional molecular

15   simulation software operable to retrieve co-ordinate data from said memory

means and to display a three-dimensional representation of a molecule on said

visual display means and being operable to produce a modified three-

dimensional homologue representation responsive to operator-selected changes

to the structure of the Aurora kinase and to display the three-dimensional

20   representation of the modified three-dimensional homologue;

(iii)   inputting three-dimensional co-ordinate data of atoms of Aurora kinase into the

computer and storing said data in the memory means;

(iv)   inputting into the data input means of said computer at least one operator-

selected change in structure of the Aurora kinase;

25 (v)   executing said molecular simulation software to produce a modified three-

dimensional molecular representation of the homologue structure;

(vi)   displaying the three-dimensional representation of the homologue on said

visual display means, whereby changes in three-dimensional structure of the

Aurora kinase resulting from changes on structure can be visually monitored;

30 (vii)   repeating steps (iv) through (vi) to produce a multiplicity of homologues; and

(viii) selecting a homologue structure represented by a three-dimensional

representation wherein the three-dimensional configuration and spatial

arrangements of the kinase catalytic domain remain substantially preserved,

-110-

thereby producing a homologue of Aurora kinase that mimics the three-dimensional structure of the Aurora kinase.

The present invention also relates to a method of producing a modulator of Aurora kinase (particularly inhibitors), such as a modulator of Aurora kinase A. The method comprises identifying a compound or molecule or designing a compound or molecule that fits into the active site ATP binding pocket of the Aurora kinase, wherein the ATP binding pocket is defined by (a) Arg136, Leu138, Gly139, Lys140, Gly141, Val146, Lys161, Leu163, Val177, Glu180, Val181, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278, and His279, the atomic coordinates being listed in Tables 1 and 2 or (b) the x,y,z- coordinates of atoms in the set of amino acid residues (set B) given by the list Arg136, Leu138, Gly139, Val146, Ala159, Lys161, Leu163, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260 and Leu262, each having coordinates as described in Tables 1 and 2, thereby producing a modulator of Aurora kinase.

Another object of the invention is to provide a method for determining the three-dimensional structure of the catalytic domain of an Aurora kinase protein, or the catalytic domain of an Aurora kinase protein in complex with an inhibitor, using homology modeling techniques and structural coordinates for a composition of this invention. Homology modeling involves constructing a model of an unknown structure using structural coordinates of one or more related proteins, protein domains and/or subdomains. Homology modeling may be conducted by fitting common or homologous portions of the protein or peptide whose three dimensional structure is to be solved to the three dimensional structure of homologous structural elements. This approach can be used to rebuild part or all of a three dimensional structure with replacement of amino acids (or other components) by those of the related structure to be solved. For example, using the structural coordinates of the catalytic domain of an Aurora kinase in complex with AMP-PNP or the inhibitor of formula II, it is possible to determine the three dimensional structure of the catalytic domain of another Aurora kinase protein through the use of homology modeling. Those coordinates may be stored, displayed, manipulated and otherwise used in like fashion as the Aurora kinase coordinates of Tables 1-2.

Thus, crystalline compositions of this invention provide a starting material for use in solving the three-dimensional structure of other Aurora kinase polypeptides.

-111-

By way of further example, the structure defined by the machine readable data may be computationally evaluated for its ability to associate with various chemical entities. The term "chemical entity", as used herein, refers to chemical compounds, complexes of at least two chemical compounds, and fragments of such compounds or complexes.

5   For instance, a first set of machine-readable data defining the three-dimensional structure of an Aurora kinase family protein, or a portion or complex thereof, is combined with a second set of machine-readable data defining the structure of a chemical entity or moiety of interest using a machine programmed with instructions for evaluating the ability of the chemical entity or moiety to associate with the Aurora kinase family protein or portion or

10   complex thereof and/or the location and/or orientation of such association. Such methods provide insight into the location, orientation and energetics of association of the Aurora kinase family protein with such chemical entities. Chemical entities that associate or interact with an Aurora kinase may inhibit its interaction with naturally occurring ligands for the protein and may inhibit biological functions mediated by such interaction. Such chemical entities are drug

15   candidates.

The protein structure encoded by the data may be displayed in a graphical format permitting visual inspection of the structure, as well as visual inspection of the structure's association with chemical entities. Alternatively, more quantitative or computational methods may be used. For example, one method of this invention for evaluating the ability of a

20   chemical entity to associate with any of the molecules or molecular complexes set forth herein comprises the steps of: a) employing computational means to perform a fitting operation between the chemical entity and a binding pocket or other surface feature of the molecule or molecular complex; and b) analyzing the results of the fitting operation to quantify the association between the chemical entity and the binding pocket.

25   This invention further provides for the use of the structural coordinates of a crystalline composition of this invention, or portions thereof, to identify reactive amino acids, such as cysteine residues, within the three-dimensional structure, such as within or adjacent to the catalytic domain; to generate and visualize a molecular surface, such as a water-accessible surface or a surface comprising the space-filling van der Waals surface of all atoms; to

30   calculate and visualize the size and shape of surface features of the protein or complex, e.g., ligand binding pockets; to locate potential H-bond donors and acceptors within the three-dimensional structure, preferably within or adjacent to a ligand binding site; to calculate regions of hydrophobicity and hydrophilicity within the three-dimensional structure,

-112-

preferably within or adjacent to a ligand binding site; and to calculate and visualize regions on or adjacent to the protein surface of favorable interaction energies with respect to selected functional groups of interest (e.g. amino, hydroxyl, carboxyl, methylene, alkyl, alkenyl, aromatic carbon, aromatic rings, heteroaromatic rings, substituted and unsubstituted

5 phosphates, substituted and unsubstituted phosphonates, substituted and unsubstituted fluoro and difluorophosphonates; etc.). One may use the foregoing approaches for characterizing the protein and its interactions with moieties of potential ligands to design or select compounds capable of specific covalent attachment to reactive amino acids (e.g., cysteine) and to design or select compounds of complementary characteristics (e.g., size, shape, charge,

10 hydrophobicity/hydrophilicity, ability to participate in hydrogen bonding, etc.) to surface features of the protein, a set of which may be preselected. Using the structural coordinates, one may also predict or calculate the orientation, binding constant or relative affinity of a given ligand to the protein in the complexed state, and use that information to design or select compounds of improved affinity.

15   In such cases, the structural coordinates of the Aurora kinase family protein, or portion or complex thereof, are entered in machine readable form into a machine programmed with instructions for carrying out the desired operation and containing any necessary additional data (e.g. data defining structural and/or functional characteristics of a potential ligand or moiety thereof, defining molecular characteristics of the various amino acids).

20   One method of this invention provides for selecting from a database of chemical structures a molecular compound capable of binding to an Aurora kinase family protein (e.g., coordinates defining the three dimensional structure of an Aurora kinase family protein or a portion thereof). Points associated with the three dimensional structure (structural coordinates) of a crystalline form of Aurora A kinase catalytic domain are characterized with

25 respect to the favorability of interactions with one or more functional groups. A database of chemical structures is then searched for candidate compounds containing one or more functional groups disposed for favorable interaction with the protein based on the prior characterization. Compounds having structures which best fit the points of favorable interaction with the three dimensional structure are thus identified.

30   It is often preferred, although not required, that such searching be conducted with the aid of a computer. In that case a first set of machine-readable data defining the three-dimensional structure of an Aurora kinase family protein, or a portion or complex thereof, is combined with a second set of machine readable data defining one or more moieties or

-113-

functional groups of interest, using a machine programmed with instructions for identifying preferred locations for favorable interaction between the functional group(s) and atoms of the protein. A third set of data, which defines the location(s) of favorable interaction between protein and functional group(s) is generated. The third set of data is then combined with a

5 fourth set of data defining the three-dimensional structures of one or more chemical entities using a machine programmed with instructions for identifying chemical entities containing functional groups to best fit the locations of their respective favorable interaction with the protein.

Compounds of the structures selected or designed by any of the foregoing means may

10 be tested for their ability to bind to an Aurora kinase family protein, inhibit the binding of an Aurora kinase family protein to a natural or non-natural ligand therefor, and/or inhibit a biological function mediated by an Aurora kinase family member.

The new crystal may be a crystal of a homologue, variant, mutant, or inhibitor complex of Aurora. The shape of the Aurora active site binding pocket in the new crystal

15 model is an equivalent shape to that of the first. The active site binding pocket of the original Aurora A crystal is defined by the amino acid residues of set A and their atomic coordinates as given in Tables 1 and 2. Equivalent shape is defined as having an rmsd of less than 1 Å upon superposition of the subset of least flexible amino acid residues (set B).

Thus, the invention provides a method to determine or design the three dimensional

20 structure of a crystal form of Aurora by difference Fourier or Molecular Replacement, using the coordinates (Tables 1 and 2) of an Aurora A crystal to model the structure of a new Aurora crystal wherein the active site ATP binding region is equivalent to that in the first crystal. The method may be carried out as follows. An Aurora protein (wild type, mutant, variant or homologue) is purified and crystallised as a pure protein or in complex with an inhibitor

25 compound. This crystal may have the same crystal form (same protein packing) as one of the crystal structures defined by Tables 1 and 2, or it may have a different crystal form (different protein packing). By taking diffraction measurements of the crystal and using the atomic coordinates in Tables 1 or 2 (or equivalent coordinates), it is possible to work out the structure of the crystal by the known methods of difference Fourier (same packing) or molecular

30 replacement (different packing). This invention covers the use in drug design of the active site ATP binding pocket in any new crystal since this will be equivalent to that in the original crystal.

-114-

The invention further provides Aurora A proteins (including homologues, variants and mutants) designed by the above method. The Aurora A proteins may have identical properties to wild type Aurora A or may have one or more different properties compared to wild type Aurora A.

5   According to a further aspect of the invention, we provide a method to select or design chemical modulators (preferably inhibitors) of Aurora by using the Aurora A catalytic domain structure (including that of homologues, variants, mutants, and inhibitor complexes) and the shape of the active site ATP binding pocket (or an equivalent shape as previously defined). Information from the three dimensional atomic coordinates of the AMP-PNP molecule and its

10   spatial orientation in relation to the three dimensional atomic coordinates of the Aurora A catalytic domain is used as a tool to design Aurora modulators (preferably inhibitors). In addition, information from the three dimensional atomic coordinates of the inhibitor molecule of formula II and its spatial orientation in relation to the three dimensional atomic coordinates of the Aurora A catalytic domain is used as tool to design Aurora modulators (preferably

15   inhibitors). Small-molecule modulators of Aurora may be selected or designed to fit into the shape of the active site binding pocket.

Knowledge of the structural determinants that account for the difference in substrate specificity between Aurora A and other kinases, such as PKA, provides a foundation for the design of highly specific modulators of the Aurora A enzyme. Structural differences at the

20   ATP binding pocket between Aurora and other kinases (defined by differences in the atomic coordinates of residues in the ATP pocket) may be used to design selective Aurora A modulators.

According to a further aspect of the invention, use of the coordinates of the Aurora A catalytic domain (Tables 1 and 2) to locate other pockets for interaction by small molecule

25   modulators that affect Aurora activity is claimed. Such pockets may overlap with the active site ATP binding pocket or be completely independent. The three-dimensional structure of Aurora A kinase is an essential tool in the discovery of any such pockets that provide an alternative for modulator interaction to the active site ATP binding pocket.

As described above, the Aurora A crystal structure may be used in the rational design

30   of drugs which modulate (preferably inhibit) the action of Aurora. These Aurora modulators may be used to prevent or treat the undesirable physical and pharmacological consequences of inappropriate Aurora activity.

-115-

The present invention will now be described with reference to the following non-limiting Examples.

Definition of Terms

5   In the Description (including the Examples) the following terms are used:

The term "atomic co-ordinates" refers to mathematical co-ordinates corresponding to the positions of every atom derived from mathematical equations related to the diffraction patterns obtained from a monochromatic beam of X-rays illuminating a crystal. The diffraction data are used to calculate an electron density map of the repeating unit of the

10 crystal. The electron density maps are used to establish the positions of the individual atoms within the unit cell of the crystal. Those of skill in the art understand that a set of atomic co-ordinates determined by X-ray crystallography is not without standard error or experimental variation.

The term "unit cell" refers to the basic building block from which the entire volume of

15 a crystal may be constructed.

The term "space group" refers to the arrangement of symmetry elements within a unit cell.

The term "molecular replacement" refers to a method that involves generating a preliminary model of a crystal whose atomic co-ordinates are not known, by orienting and

20 positioning a related molecule whose atomic co-ordinates are known. Phases are then calculated from this model and combined with observed amplitudes to give an approximate Fourier synthesis of the structure whose co-ordinates are unknown.

Example 1: Production of the kinase catalytic domain of Aurora A

25 Molecular Biology:

In order to obtain a polypeptide (or protein) that can be utilised for determination of the three dimensional (tertiary) structure of Aurora A, DNA encoding Aurora A may be obtained by total gene synthesis or by cloning. This DNA may then be expressed in a suitable expression system to obtain a polypeptide that can be subjected to techniques to determine its

30 three dimensional structure.

In this case, the human Aurora A gene carrying an artificially induced mutation (GAT to ACT in nucleotides 862-864, taking the A of the initial ATG in the open reading frame of the gene to be +1) encoding for a threonine to aspartate (T to D using the single letter amino

-116-

acid code) mutation of amino acid 287 (taking the first amino acid immediately after the initial methionine as amino acid number one) formed the basis of the expression construct used in these studies. This [T287D]Aurora A mutant was a gift of Dr. Jim Bischoff, SUGEN Inc. Since the full length [T287D]Aurora A protein expressed in *E.coli* was poorly soluble

5 and aggregated on purification, a truncated mutant form was generated. The regions encoding for amino acids 94 to the stop codon of [T287D] Aurora A was amplified using the polymerase chain reaction (PCR). The 5' PCR primer (5'GATCGATCGGATCCACCCAAAAGAGCAAGCAGCCC 3'; SEQ ID NO.: 1) carried a spacer region (to allow efficient cleavage by restriction endonuclease), the BamH1restriction

10 endonuclease recognition sequence and sequence corresponding to the bases 283-301. The 3' primer (5' TGACGCTAGGATCCCCTAAGACTGTTTGCTAGCTGATTC 3'; SEQ ID NO.: 2) carried a spacer region, BamH1 recognition sequence and 3' end of the Aurora A (bases 1189-1212) sequence including the stop codon. PCR products were purified and cloned in to the pCR-Script vector (Stratagene) using the pCR-Script AMP cloning kit (Stratagene.

15 product # 211188) according to manufacturers directions. The pCR-Script vector carrying the [T287D] Aurora A (94-402) sequence was digested with BamH1, the digestion products resolved by agarose gel electrophoresis and the DNA fragment corresponding to the [T287D] Aurora A (94-402) sequence excised and purified using a Qiagen QIAquick kit (Qiagen product #28704). This fragment was then ligated into the vector pTB375NBSE, which had

20 previously been cut with BamH1. (Details of the methods for the assembly of recombinant DNA molecules can be found in standard texts, for example Sambrook et al. 1989, Molecular Cloning – A Laboratory Manual, 2nd Edition, Cold Spring hbor Laboratory Press and Ausubel et al. 1999, Current Protocols in Molecular Biology, John Wiley and Sons Inc). The pTB375NBSE vector is derived from pAT153, which is a mobilization-minus derivative of

25 pBR322. The inserted genes were under the control of a bacteriophage T7 promoter and therefore requires expression of the T7 polymerase in *trans* for efficient transcription in *E.coli*. The plasmid encodes tetracyline resistance for selection.

The ligation reactions were transfected in-to TOP10 competent *E.coli* (Invitrogen product #C4040-10) and *E.coli* carrying the pTB375NBSE recombinant vectors identified by

30 their ability to grow on media containing tetracycline. Plasmid DNA was extracted from these bacteria and subjected to digestion with the restriction endonuclease EcoR1 to identify those carrying the [T287D] Aurora A (94-402) sequence. The identity of the insert was then confirmed by dideoxy chain termination DNA sequencing prior to protein expression.

-117-

pTB375NBSE carries the initiation codon (ATG) 3' to the T7 promoter and also the following sequence up to and including the BamH1 restriction endonuclease recognition site:

5'... ATG GGC CAT CAT CAT CAT CAT CAC GGA TCC ......3' (SEQ ID NO.: 3)

5

Sequences inserted into the BamH1 site "in frame" with the initiation codon will therefore be expressed as a fusion protein with the following N-terminal fusion:

(N-terminal) MGHHHHHHGS......(C-terminal) (SEQ ID NO.: 4)

10

The fusion of 6 histidines to proteins is commonly used to provide a "tag" for protein purification, usually by affinity for metal ions such as nickel. Since the [T287D] Aurora A sequence coding for amino acids 94-402 was inserted in to the BamH1 site, the plasmid encodes for the following protein (using the standard single letter amino acid code):

15

MGHHHHHHGSTQKSKQPLPSAPENNPEEELASKQKNEESKKRQWALEDFEIGRPLGK
GKFGNVYLAREKQSKFILALKVLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYGY
FHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIK
PENLLLGSAGELKIADFGWSVHAPSSRRTDLCGTLDYLPPEMIEGRMHDEKVDLWSL

20 GVLCYEFLVGKPPFEANTYQETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPML
REVLEHPWITANSSKPSNCQNKESASKQS. (SEQ ID NO.: 5)

This protein will be referred to in the text as MG-6His-GS-[T287D]Aurora A(94-402).

25 Based on the limited proteolysis studies (described later in Example 1), two additional truncated mutant forms of the Aurora A protein were also generated. The regions encoding for amino acids 113-400 and 122-400 of [T287D]Aurora A were amplified using the polymerase chain reaction (PCR). The 5' PCR primers (5'CATATGCTGGCATCAAAACAGAAAATG 3' for 113-400 of [T287D]Aurora A or 5'CATATGTCAAAAAAGAGGCAGTGGGC 3' for

30 122-400 of [T287D]Aurora A) carried a Nde1 restriction endonuclease recognition sequence. A single 3' primer (5'GGATCCTCATTTGCTAGCTGATTCTTTGTTTTGG 3') was used for both constructs and carries a BamH1 recognition sequence and 3' end of the Aurora A sequence following the stop codon. PCR products were purified and cloned into the pCR-

-118-

Script vector (Stratagene) using the pCR-Script AMP cloning kit (Stratagene. product #

211188) according to manufacturers instructions and transfected into the *E. coli* strain DH5α

(Invitrogen product #18258-012). The *E. coli* colonies containing the recombinant pPCR-

Script [T287D]Aurora A(113-400) or pPCRscript [T287D]Aurora A(122-400) were

5   identified by colony PCR screening using the primers T3

(5'AATTAACCCTCACTAAAGGG 3') and T7pro (5'TAATACGACTCACTATAGGG 3')

hybridising specifically on either side of the pPCR script vector cloning site. The pPCR-Script

vectors carrying the [T287D]Aurora A(113-400) or [T287D]Aurora A(122-400) sequence

were prepared from *E. coli* and were digested with Nde 1 and Bam H1, the digestion products

10 resolved by agarose gel electrophoresis. The fragments containing the [T287D]Aurora A(113-

400) or [T287D]Aurora(122-400) were ligated into the expression vector pET28a (Novagen

product #69864-3) between the Nde1 and BamH1 restriction sites. The inserted genes were

cloned in frame with a sequence coding for a 6 histidine tag followed by a sequence encoding

a thrombin protease cleavage site (see below for a complete description). The inserted genes

15 are under the control of a bacteriophage T7 promoter and therefore require expression of the

T7 polymerase in *trans* for efficient transcription in *E.coli*. The plasmid encodes kanamycin

resistance for selection.

The ligation reactions were transfected into DH5α competent *E.coli* and the bacteria

carrying the pET28a-[T287D]Aurora A(113-400) or pET28a-[T287D]Aurora A(122-400)

20 recombinant vectors were identified by their ability to grow on media containing kanamycin.

Plasmid DNAs were extracted from these bacteria and subjected to digestion with the

restriction endonucleases Nde1 and BamH1 to identify those carrying the [T287D]Aurora

A(113-400) or [T287D]Aurora A(122-400) sequences. The identity of the insert was then

confirmed by dideoxy chain termination DNA sequencing prior to protein expression.

25   pET28a carries the initiation codon (ATG) 3' to the T7 promoter and also the

following sequence up to and including the Nde1 restriction endonuclease recognition site:

5'... ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG

GTG CCG CGC GGC AGC CAT ATG ......3'

30

-119-

Sequences inserted into the Nde1 site "in frame" with the initiation codon will therefore be expressed as a fusion protein with the following N-terminal fusion (using the standard single letter amino acid code):

5   (N-terminal) MGSSHHHHHHSSGLVPRGSHM......(C-terminal)

The fusion of 6 histidines to proteins is commonly used to provide a "tag" for protein purification, usually by affinity for metal ions such as nickel. The motif "LVPRGS" corresponds to a specific thrombin protease cleavage site that allows the proteolytic removal

10 of the sequence "MGSSHHHHHHSSGLVPR" after incubation of the protein with thrombin. Since the [T287D]Aurora A sequences coding for amino acids 113-400 and 122-400 were inserted into the Nde 1 site, the plasmid encodes for the following protein (using the standard single letter amino acid code):

15 [T287D]Aurora A(113-400)

MGSSHHHHHHSSGLVPRGSHMLASKQKNEESKKRQWALEDFEIGRPLGKGKFGNVY
LAREKQSKFILALKVLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYGYFHDATRV
YLILEYAPLGTVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLG
SAGELKIADFGWSVHAPSSRRTDLCGTLDYLPPEMIEGRMHDEKVDLWSLGVLCYEF

20 LVGKPPFEANTYQETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPMLREVLEHP
WITANSSKPSNCQNKESASK

This protein will be referred to in the text as MGSS-6His-SSGLVPRGSHM-[T287D]Aurora A(113-400)

25 [T287D]Aurora A(122-400)

MGSSHHHHHHSSGLVPRGSHMSKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFI
LALKVLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLG
TVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLGSAGELKIADF
GWSVHAPSSRRTDLCGTLDYLPPEMIEGRMHDEKVDLWSLGVLCYEFLVGKPPFEAN

30 TYQETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPMLREVLEHPWITANSSKPS
NCQNKESASK

This protein will be referred to in the text as MGSS-6His-SSGLVPRGSHM-[T287D]Aurora A(122-400)

-120-

Protein expression

pTB375NBSE carrying the [T287D] Aurora A (94-402) sequence were transfected

into *E.coli* BL21(DE3) pLys S (genotype: B F dcm $ompT$ $hsdS$(r_B^- m_B^-) gal λ(DE3) [pLysS

5 Cam^r]). The strain was grown for 16 h (LB medium containing tetracycline (10μg/mL) and

chloramphenicol (34μg/mL) at 30°C in shake flasks to OD$_{550nm}$ ~5. This culture was

inoculated into high biomass medium containing tetracycline (10μg/mL) and chloramphenicol

(34μg/mL), in a 20 L fermenter (B. Braun, Melsungen, Germany). Cells were grown

aerobically in fed batch culture at 30 °C, pH 6.7 with dissolved oxygen tension maintained at

10 50% air saturation. Expression of 6His-[T287D] Aurora A (94-402) was induced 12 hours

post inoculation (OD$_{550nm}$ ~13) with 0.40mM isopropyl-β-D-thiogalactopyranoside (IPTG),

and cells harvested 3.0 hours later (OD$_{550nm}$ ~33) by batch centrifugation (7,000xg at 4 °C for

30 min).

15   pET28a carrying the [T287D]Aurora A(122-400) sequence was transfected into *E.coli*

DS410 (DE3) (a derivative of the original minicell-producing strain P678-54). The strain was

grown for 30 h (M9 glucose medium containing kanamycin (25μg/mL) at 37°C in shake

flasks to OD$_{550nm}$ ~1.4. This culture was inoculated into high biomass medium containing

kanamycin (25μg/mL) in a 20 L fermenter (B. Braun, Melsungen, Germany). Cells were

20 grown aerobically in fed batch culture at 30°C, pH 6.7 with dissolved oxygen tension

maintained at 50% air saturation. Expression of MGSS-6His-SSGLVPRGSHM-

[T287D]Aurora A(122-400) was induced 16 hours post inoculation (OD$_{550nm}$ ~19) with

0.10mM isopropyl-β-D-thiogalactopyranoside (IPTG), and cells harvested 23 hours later

(OD$_{550nm}$ ~26) by batch centrifugation (7,000xg at 4°C for 30 min).

25

Definition of kinase domain fragment

MG-6His-GS-[T287D]Aurora A(94-402) was purified from *E. coli* cell paste by Ni-

NTA Agarose chromatography followed by size exclusion chromatography. The solution

30 properties of this protein were found to be unfavourable for structural studies. Limited

proteolysis of MG-6His-GS-[T287D]Aurora A(94-402) was used to identify fragments of the

Aurora A kinase domain with superior solution properties. Aliquots of MG-6His-GS-

[T287D]Aurora A(94-402) were subjected to proteolytic digestion with trypsin, thermolysin

-121-

or endoproteinase Glu-C (V8). Protein fragments were identified by analysis with Coomassie-stained SDS PAGE, electrospray mass spectrometry (ESMS) and N-terminal sequencing. Cleavage and purification was performed at sufficient scale to produce appropriate quantities of [T287D]Aurora A (122-396) for crystallisation, as detailed below. Characterization of

5 these fragments of Aurora A was used to design additional constructs, including MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400). The molecular biology procedures used to generate MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400) are given in the Molecular Biology section of Example 1.

10 Lysis of E. coli containing MG-6His-GS-[T287D]Aurora A(94-402)

The following procedures were performed at 4°C unless otherwise stated.

E. coli cell paste (200 g) was resuspended using a Kinematica PT6000 homogeniser (Kinematica GMBH, Basel, Switzerland) in 1.0 l of lysis buffer (40mM HEPES, 200mM NaCl, 2mM imidazole, 2mM 2-mercaptoethanol, 1mM benzamidine, pH 7.4). The cells were

15 lysed using an Avestin EmulsiFlex ñC5 (Avestin, Inc., Ottawa, Canada), using a single pass at an average pressure of 10,000 psi. The resulting lysate was centrifuged at 17,000 x g (average) for 90 min before aspirating the supernatant and discarding the pellet.

Lysis of E. coli containing MG-6His-GS-[T287D]Aurora A(94-402)

20 The following procedures were performed at 4 °C unless otherwise stated.

E. coli cell paste (200 g) was resuspended using a Kinematica PT6000 homogeniser (Kinematica GMBH, Basel, Switzerland) in 1.0 l of lysis buffer (40mM HEPES, 200mM NaCl, 2mM imidazole, 2mM 2-mercaptoethanol, 1mM benzamidine, pH 7.4). The cells were lysed using an Avestin EmulsiFlex ñC5 (Avestin, Inc., Ottawa, Canada), using a single pass at

25 an average pressure of 10,000 psi. The resulting lysate was centrifuged at 17,000 x g (average) for 90 min before aspirating the supernatant and discarding the pellet.

Preparation of [T287D]Aurora A(122-396)

The following procedures were performed at 4°C unless otherwise stated.

30 A 26 mm diameter chromatography column packed with 25 mL Qiagen Ni NTA-Agarose (Qiagen GMBH, Hilden, Germany) was equilibrated with 10 column volumes of lysis buffer before loading lysate supernatant containing MG-6His-GS-[T287D]Aurora A(94-402) onto

-122-

the column at a flow rate of 0.9 mL/min. Using a flow rate of 2.0 mL/min the column was

washed with 10 column volumes of wash buffer (40mM HEPES, 20mM imidazole, 2mM 2-

mercaptoethanol, pH 7.5) to remove weakly bound or non-specifically bound impurities.

Elution of bound protein was effected using elution buffer (40mM HEPES, 400mM

5   imidazole, 2mM 2-mercaptoethanol, pH 7.5) at 2.0 mL/min. Eluted material was flowed

through a second chromatography column (26 mm diameter, packed with 25 mL Pharmacia Q

Sepharose Fast Flow (Amersham Pharmacia Biotech, Uppsala, Sweden) previously

equilibrated with 10 column volumes of wash buffer). Fractions of 10.0 mL were collected,

and after analysis by Coomassie-stained SDS PAGE, those fractions containing significant

10   amounts of MG-6His-GS-[T287D]Aurora A(94-402) were pooled. At this stage the pool

(approximately 200 mL) was stored in an airtight container at 4 °C for up to seven days.

From this stage forward, all procedures were carried out at room temperature, unless

otherwise stated. A Pharmacia HiPrep 16/60 Sephacryl S-100 pre-packed size exclusion

column was equilibrated in running buffer (40 mM HEPES pH7.5, 350 mM NaCl, 2 mM

15   dithiothreitol (DTT)). The column was run at a flowrate of 1.0 mL/min. A 10 mL sample of

the MG-6His-GS-[T287D]Aurora A(94-402) pool was centrifuged (31,000 x g, 4 °C, 60 min)

and loaded onto the column. The fractions (2.0 mL) were analysed by Coomassie-stained SDS

PAGE, and those containing MG-6His-GS-[T287D]Aurora A(94-402) were pooled.

Limited proteolysis at room temperature was carried out on the size exclusion

20   chromatography-purified pool of MG-6His-GS-[T287D]Aurora A(94-402), whose

concentration was 1 mg/mL. Using a mass ratio of 1 part protease to 100 parts MG-6His-GS-

[T287D]Aurora A(94-402), endoproteinase Glu-C from *Staphylococcus aureus* V8

(Boehringer Mannheim UK, Lewes, Sussex, UK) was added to the pool. Proteolysis was

allowed to continue for between 3 and 7 h.

25   A chromatography column was packed with a mixture of Pharmacia Sephacryl S-100

HR and Pharmacia Q-Sepharose high performance in the ratio of 9:1 v/v respectively (referred

to as 'the S-100/Q column'). The column volume was 130 mL. It was equilibrated and run in

S-100/Q running buffer (40 mM HEPES pH7.5, 50 mM NaCl, 2 mM dithiothreitol) at a

flowrate of 1.0 mL/min. A sample of the proteolysed pool (8 mL) was loaded onto the S-

30   100/Q column and 2.0 mL fractions were collected. Fractions were analysed by Coomassie-

stained SDS PAGE, and those containing significant quantities of pure [T287D]Aurora

A(122-396) were pooled. A sample of the pool was analysed by LC-ESMS using a Micromass

-123-

LCT in conjunction with a Waters Alliance HPLC (Micromass, Manchester, UK). A further sample of the pool was subjected to N-terminal sequencing. Once the identity of the cleaved protein had been confirmed as [T287D]Aurora A(122-396), it was submitted for crystallisation.

5

Preparation of GSHM-[T287D]Aurora A(122-400)

The following procedures were performed at 4°C unless otherwise stated.

A 26 mm diameter chromatography column packed with 15 mL Qiagen Ni NTA-Agarose (Qiagen GMBH, Hilden, Germany) was equilibrated with 10 column volumes of lysis buffer

10 before loading lysate supernatant containing MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400) onto the column at a flow rate of 1.0 mL/min. Using a flow rate of 2.0 mL/min the column was washed with 7 column volumes of wash buffer (40mM HEPES, 200mM NaCl, 10mM $MgCl_2$, 20mM imidazole, 2mM 2-mercaptoethanol, pH 7.5) to remove weakly bound or non-specifically bound impurities. Elution of bound protein was effected using

15 elution buffer (40mM HEPES, 400mM imidazole, 10mM $MgCl_2$, 2mM 2-mercaptoethanol, pH 7.5) at 2.0 mL/min. Fractions of 10.0 mL were collected, and after analysis by Coomassie-stained SDS PAGE, those fractions containing significant amounts of MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400) were pooled.

From this stage in the purification onward, all procedures were carried out at room

20 temperature unless otherwise stated. A Pharmacia HiPrep 26/10 Fast Desalting pre-packed column was equilibrated in running buffer (40 mM HEPES pH7.4, 150 mM NaCl, 2 mM 2-mercaptoethanol). The column was run at a flowrate of 3.0 mL/min. A 10 mL sample of the MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400) pool was filtered (0.22μm) and loaded onto the column. Fractions were collected, and those containing the most concentrated

25 amounts of MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400) were pooled.

Bovine thrombin (500 units, Amersham Pharmacia Biotech) was added to the pool of 50 mg (+/- 20%) of purified MGSS-6His-SSGLVPRGSHM-[T287D]Aurora-A(122-400) whose concentration was 1 mg/mL. Specific proteolytic cleavage was allowed to proceed to completion at 4°C, producing the truncated mutant GSHM-[T287D]Aurora A(122-400).

30   A Pharmacia HiPrep 16/60 Sephacryl S-100 pre-packed size exclusion column was equilibrated in running buffer (40 mM HEPES pH7.4, 50 mM NaCl, 1 mM dithiothreitol). The column was run at a flowrate of 1.0 mL/min. A 10 mL sample of the GSHM-

-124-

[T287D]Aurora A(122-400) pool was filtered (0.22μm) and loaded onto the column.

Fractions (2.0 mL) were analysed by Coomassie-stained SDS PAGE, and those containing

GSHM-[T287D]Aurora A(122-400) were pooled, and submitted for crystallisation.

5 *Analysis*

For SDS PAGE all samples were diluted in Laemmli buffer containing 2-mercaptoethanol,

boiled for 2 minutes and loaded onto a 8-16% gradient, 1.5mm thickness x 10 well NOVEX

gel (NOVEX, San Diego, California). Gels were stained with Coomassie blue R-250. Edman

degradation was carried out on a Perkin Elmer 477A peptide sequencer (Applied Biosystems,

10 Foster City, CA) with on-line detection of PTH amino acids. Mass spectra were acquired

using a Micromass LCT with electrospray source (Micromass, Manchester, UK) and on-line

Waters 2790 Alliance delivery system (Waters, Milford, MA). Protein was loaded directly on

to a Phenomenex Jupiter 5μ C5 300_ 150 x 2.00 mm reverse phase column equilibrated in

Milli Q water (Millipore, Bedford, MA), 2.7% acetonitrile, 0.1% trifluoroacetic acid, and the

15 column was developed with a 2.7% to 90% acetonitrile gradient over 30 minutes at a flowrate

of 80μl/min. A fraction (approximately 25%) of the eluted proteins passed into the mass

spectrometer.

20 **Example 2: Crystallisation of [T287D] Aurora A catalytic domain constructs**

The [T287D] Aurora A(122-396):AMPPNP complex was crystallized at 15°C by the

method of hanging-drop vapour diffusion. The protein [T287D] Aurora A (122-396) was

concentrated to ~10 mg/mL solution (in 40mM HEPES pH 7.4, 2mM DTT, 50mM NaCl),

25 5mM AMP-PNP was then added to this solution and the complex was incubated on ice for 30

minutes. Prior to setting up crystallization trials this complex solution was microfuged for 10

minutes. The drops contained a 1:1 by volume mixture of complex solution and reservoir

buffer (0.2M K_2HPO_4, 1.6M NaH_2PO_4, 0.1M phosphate/citrate buffer pH 3.8) giving a final

4μl drop volume. The [T287D] Aurora(122-396)-AMP-PNP crystals belong to space group

30 $P3_221$ with unit cell dimensions $a = b = 86.55$ Å, $c = 78.34$ Å, and α=β=90°, γ=120°, and

contain 1 complex molecule per asymmetric unit. Before data collection, the crystals were

transferred briefly (for about 20 seconds) to a cryobuffer containing 0.2M K_2HPO_4, 1.6M

-125-

NaH₂PO₄, 0.1M phosphate citrate pH 3.8, 30% glycerol before being cooled to 100K in a nitrogen gas stream.

The GSHM-[T287D] Aurora A(122-400) complex with the chemically synthesized inhibitor of formula II was crystallized as follows. Preparation of compound of formula II is

5 described under example 19 in patent publication number WO 01/21597, publication date 29/3/01 (application number PCT/GB00/03593, international filing date 19/09/00). The compound was added at 5mM to a solution containing protein (GSHM-[T287D] Aurora A(122-400) at 10mg/ml, 40mM HEPES pH7.5, 50mM NaCl, and 1mM 2-mercaptoethanol. Drops were formed by mixing 1:1 volumes of protein complex solution and a reservoir

10 solution containing 22% PEG 4000 and 0.2M ammonium sulphate. Crystallisation was achieved by hanging drop vapour diffusion at 15°C. Data were collected at room temperature from a crystal mounted in a capillary. The crystal could be translated in the X-ray beam to allow multiple exposures. The Aurora A-inhibitor crystals are of space group P2₁ with unit cell dimensions a = 52.6, b = 88.4, c = 67.8 Å, α = γ = 90 and β = 90.01°, and contain two

15 complex molecules in the asymmetric unit.

Example 3: Structure determination of [T287D] Aurora A catalytic constructs

Diffraction data were collected at beamline PX9.6 at the SRS, Daresbury on an ADSC

20 Quantum 4 CCD detector. The data were indexed and integrated with the program Mosflm and scaled with the program SCALA (CCP4). Molecular replacement and rigid body refinement to a resolution of 3.0 Å were carried out using the program AMoRe. A search model was derived from mouse PKA, truncating the model at residues 32 to 310 and replacing all non-identical residues with Ala. 5% of the data were reserved at this stage as a cross-

25 validation set and the initial model underwent torsion angle simulated annealing in the program CNX using a maximum likelihood target and an overall anisotropic temperature factor correction. The model then underwent iterative rounds of manual rebuilding and simulated annealing until the working R-factor fell below 30%, at which point restrained isotropic individual temperature factor refinement was carried out. Concurrent building of

30 both inhibitor complexes with the same Aurora protein in different crystal forms proved very instructive when it came to clarification of regions that were difficult to interpret. Further iterative rebuilding and addition of waters was carried out until the free R factor converged.

Crystallographic data and refinement statistics are given in tables 4 and 5.

Table 4: Aurora-AMPPNP complex data and refinement statistics.

Space Group	$P3_221$
Cell constants	a=b=86.55, c= 78.34Å α=β=90, γ=120°
Reflections	62278
Independent Reflections	17003
R_{sym} (2.25-2.2 Å)	3.6% (32.6%)
Resolution (Å)	38-2.2
I/sigI (2.25-2.2 Å)	19.3 (3.0)
Completeness (2.25-2.2 Å)	97.1% (83.2%)
$R_{(free)}$, $R_{(work)}$	23%, 28%
Rmsd (bond lengths)	0.006
Rmsd (bond angles)	1.2

5

Table 5 : Aurora-inhibitor complex data and refinement statistics:

Space Group	$P2_1$
Cell constants	a=52.6, b=88.4, c=67.8Å α=γ=90, β=90.01°
Reflections	36664
Independent Reflections	26294
R_{sym} (2.25-2.1 Å)	6.6% (30.5%)
Resolution (Å)	52-2.1
I/sigI (2.25-2.1 Å)	7 (2.1)
Completeness (2.25-2.1 Å)	72.5% (25.5%)
$R_{(free)}$, $R_{(work)}$	22%, 27%
Rmsd (bond lengths)	0.019
Rmsd (bond angles)	1.8

Example 4: Description of the Structure of Aurora A kinase

5   The structure of [T287D] Aurora A (122-396) in a binary complex with the ATP
analogue AMP-PNP has been solved to a resolution of 2.2 Å. The structure of GSHM-
[T287D] Aurora A(122-400) in a binary complex with the synthetic inhibitor of formula II has
been solved to a resolution of 2.1 Å. The structures contain the residues of the kinase
catalytic domain. The kinase domain of [T287D] Aurora A shows the bilobal structure
10   characteristic of protein kinases with the ATP and inhibitor binding site situated between the
two lobes. The N-terminal domain (lobe) comprises a twisted β-sheet and a single kinked
helix. The C-terminal lobe comprises mainly helices but also includes a small region of β-
sheet. Parts of the polypeptide chain are disordered. In particular, the activation loop,
residues 279 to 290 containing the T287D substitution, is not visible in the electron density.
15   The disordered nature of the activation loop is a common feature in kinase crystal structures.

The structure adopts a conformation typical of catalytically inactive kinases, despite
the introduction of the constitutively active mutation, T287D. It is thought that the acidic pH
at which the crystallisation experiments were carried out will result in the introduced aspartate
being protonated, and thus no longer able to mimic the phosphorylated threonine in the wild- ·
20   type activated protein. The kinase activity of the mutant enzyme towards a peptide substrate
was measured at varying pH values, as shown in Figure 3, and indeed, activity is significantly
reduced falls as the pH is lowered.

The inactive conformation seen in our [T287D] Aurora A complexes is clearly capable
of binding the inhibitor of formula II and the ATP analogue, and therefore allows structure-
25   based design, which needs to make allowances for the flexibility and conformational changes
that the kinase may undergo, for example, between its active and inactive states. In the case of
the inhibitor, the inactive conformation may be forced by the steric bulk of the inhibitor.

Aurora A is quite closely related to the cyclic AMP-dependent protein kinase, also
known as PKA, and the structures superpose with an overall rmsd of 1.4Å. The ATP binding
30   cleft of Aurora A is more extended than the equivalent cleft in PKA on account of a shift in
the position of a helix, formed by residues 174 to 182 in the N-terminal lobe. In the structure
of [T287D] Aurora A, the helix is displaced approximately 3Å away from the ATP binding
pocket compared with the equivalent helix in PKA, thus extending the length of the cleft

-128-

between the two lobes. The extended cleft can be exploited by elongated inhibitor molecules
such as that of formula II and may be key to the design of specific inhibitors. The conserved
DFG motif (Asp273Phe274Glu275) preceding the activation loop is apparent in the electron
density. This region contains an aspartate residue necessary for catalysis. The glycine-rich
5 loop, which is important for ATP binding in all kinases, shows good electron density
throughout the main chain atoms, although the temperature factors are quite high, indicating
significant mobility of the loop. However, the density for the side chains of some residues,
such as Phe 143, is poor, and these are likely to adopt multiple conformations.

The AMP-PNP molecule adopts a dual conformation (Fig. 1). The adenine ring and
10 ribose moiety in both conformations occupy similar locations with respect to the kinase
molecule. Classical hydrogen-bonding interactions are made between the adenine ring and the
hinge region of Aurora A. These are between N6 of the adenine ring and the main chain
oxygen of Glu 210 and between N1 of the adenine ring and the main chain nitrogen of Ala
212. The differences in the two conformations arise from torsion angle differences between
15 the ribose ring and the phosphate groups and also in torsion angles of phosphorus-oxygen
bonds. In one conformation, the β-phosphate group forms a hydrogen bond to a water
molecule, which, in turn, forms a hydrogen bond to Asp 273. In the other conformation, the
β-phosphate forms hydrogen bonds with Ser 277 and Gln 260. In both conformations the α-
phosphate forms a salt-bridge with Lys 161, and also forms a hydrogen bond with the main
20 chain nitrogen of Val 278. No electron density for the γ-phosphate is present in either
conformation suggesting a high degree of disorder. This disorder of the γ-phosphate has also
been seen in other crystal structures, for example that of Checkpoint kinase.

The molecule of formula II also binds in the ATP binding site in the cleft between the
two domains in the Aurora A kinase molecule. The molecule of formula II adopts an extended
25 conformation, which demonstrates the extent of the available binding pocket (Fig 2b). A
classical kinase (adenine-mimetic) inhibitor hydrogen bond interaction with the main chain
peptides is made between N(17) in the inhibitor and the amide of amino acid residue 212. The
piperidine moiety of the inhibitor extends into solvent (on the left in Fig 2b). At the other
extreme of the inhibitor (right end in Fig 2b) the benzoyl moiety fits into a hydrophobic
30 pocket formed by residues Leu163, Leu181, Leu195, Leu207 and Trp276. This inhibitor
represents a more interesting start point for design than AMPPNP since protein regions more
remote from the ATP location are explored, and this may help achieve specificity.

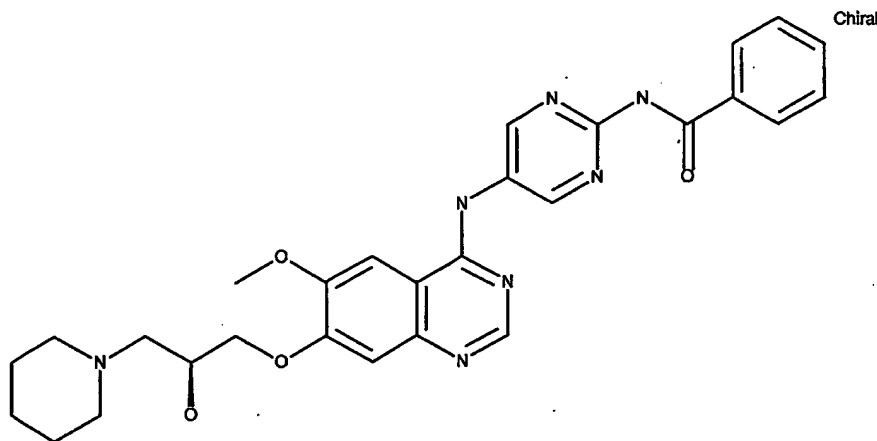## CLAIMS

What we claim is:

5   1.   A crystalline form of a polypeptide comprising the catalytic domain of Aurora kinase.

2.   A crystalline form according to Claim 1, wherein the polypeptide is an Aurora A kinase.

10 3.   A crystalline form according to Claim 1 or Claim 2, wherein the crystalline form has the space group P3$_2$21 or the space group P2.

4.   A crystalline form according to any one of the preceding claims, wherein the crystalline form has unit cell dimensions a=b=86.55, c= 78.34 Å, α=β=90 and γ=120° or unit

15 cell dimensions a = 52.6, b = 88.4, c = 67.8 Å, α= γ = 90 and β= 90.01 °.

5.   A crystalline form according to any one of the preceding claims, wherein the catalytic domain comprises a binding site, wherein the binding site is defined by the x,y,z-coordinates of atoms in the set of amino acid residues given by the list: Arg136, Leu138, Gly139, Lys140,

20 Gly141, Val146, Ala159, Lys161, Leu163, Val177, Glu180, Val181, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278, and His279 or their equivalent, wherein the atomic coordinates are listed in Tables 1 and 2; or wherein the binding site is defined by the x,y,z-co-ordinates of atoms in the

25 set of amino acid residues given by the list: Arg136, Leu138, Gly139, Val146, Ala159, Lys161, Leu163, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260 and Leu262 or their equivalent, and wherein the x,y,z-coordinates are within a root mean square deviation of not more than 1.0 Å of the coordinates listed in Tables 1 and 2.

30
6.   A crystalline form according to any one of the preceding claims, additionally comprising an Aurora kinase inhibitor in complex with the catalytic domain of Aurora kinase.

-130-

7.   A crystalline form according to Claim 6, wherein the Aurora kinase inhibitor is a compound of formula II:

Chiral

Formula II.

8.   A method of designing an Aurora chemical modulator using the atomic coordinates of a crystalline form according to any one of claims 1 to 5.

9.   A method of selecting an Aurora chemical modulator using the atomic coordinates of a crystalline form according to any one of claims 1 to 5.

10.   A method of designing an Aurora protein using the atomic coordinates of a crystalline form according to any one of claims 1 to 5.

11.   A method of designing or selecting an Aurora modulator comprising the steps of: (a) exploring the atomic coordinates of Aurora as presented in Table 1 and Table 2 for information on the three-dimensional characteristics of the protein surface; (b) arriving at an alternative overlapping or non-overlapping binding pocket to the active site ATP binding pocket; and (c) selecting or designing an Aurora modulator using the binding pocket information.

12.   A method of designing the three-dimensional structure of a second crystal form of Aurora kinase comprising the step of applying difference Fourier or molecular replacement methods using the atomic coordinates of an original crystal as presented in Table 1 and Table 2 to model the structure of the crystal second form, wherein the active site ATP binding pocket of the second crystal form is equivalent to that in the original crystal.

-131-

13.   A method of designing or selecting an Aurora kinase modulator using the coordinates
of any protein shown by this invention to possess structural similarity or relevance to Aurora
kinase.

14.   A method for designing a homologue of Aurora kinase that mimics the three-
dimensional structure of Aurora kinase, which comprises:

(i)   determining the three-dimensional coordinates of atoms of an Aurora kinase;

(ii)   providing a computer having a memory means, a data input means, a visual
display means, said memory means containing three-dimensional molecular
simulation software operable to retrieve co-ordinate data from said memory
means and to display a three-dimensional representation of a molecule on said
visual display means and being operable to produce a modified three-
dimensional homologue representation responsive to operator-selected changes
to the structure of the Aurora kinase and to display the three-dimensional
representation of the modified three-dimensional homologue;

(iii)   inputting three-dimensional co-ordinate data of atoms of Aurora kinase into the
computer and storing said data in the memory means;

(iv)   inputting into the data input means of said computer at least one operator-
selected change in structure of the Aurora kinase;

(v)   executing said molecular simulation software to produce a modified three-
dimensional molecular representation of the homologue structure;

(vi)   displaying the three-dimensional representation of the homologue on said
visual display means, whereby changes in three-dimensional structure of the
Aurora kinase resulting from changes on structure can be visually monitored;

(vii)   repeating steps (iv) through (vi) to produce a multiplicity of homologues;

(ix)   selecting a homologue structure represented by a three-dimensional
representation wherein the three-dimensional configuration and spatial
arrangements of the kinase catalytic domain remain substantially preserved,
thereby producing a homologue of Aurora kinase that mimics the three-
dimensional structure of the Aurora kinase.

15.   A method of producing a modulator of Aurora kinase comprising identifying a
compound or molecule or designing a compound or molecule that fits into the active site ATP

-132-

binding pocket of the Aurora kinase, wherein the ATP binding pocket is defined by the x,y,z-coordinates of atoms in the set of amino acid residues given by the list (a) Arg136, Leu138, Gly139, Lys140, Gly141, Val146, Lys161, Leu163, Val177, Glu180, Val181, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215,

5 Thr216, Arg219, Glu259, Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278, and His279, the atomic coordinates being listed in Tables 1 and 2 or (b) the x,y,z- coordinates of atoms in the set of amino acid residues given by the list Arg136, Leu138, Gly139, Val146, Ala159, Lys161, Leu163, Ile183, Gln184, Leu193, Leu195, Leu207, Leu209, Glu210, Tyr211, Ala212, Pro213, Leu214, Gly215, Thr216, Arg219, Glu259, Asn260 and

10 Leu262, each having coordinates as described in Tables 1 and 2, thereby producing a modulator of Aurora kinase.

16.   A crystalline form wherein the catalytic domain comprises a binding site , wherein the binding site is defined by the x,y,z-co-ordinates of atoms in the set of amino acid residues

15 given by the list: Leu138, Gly139, Val146, Lys161, Val177, Arg178, Arg179, Glu180, Val181, Glu182, Ile183, Gln184, Leu193, Leu209, Tyr211, Ala212, Gly215, Thr216, Glu259, Asn260, Leu262, Ala272, Asp273, Phe274, Gly275, Trp276, Ser277, Val278 and His279 or their equivalent, wherein the atomic co-ordinates are listed in Table 1a.
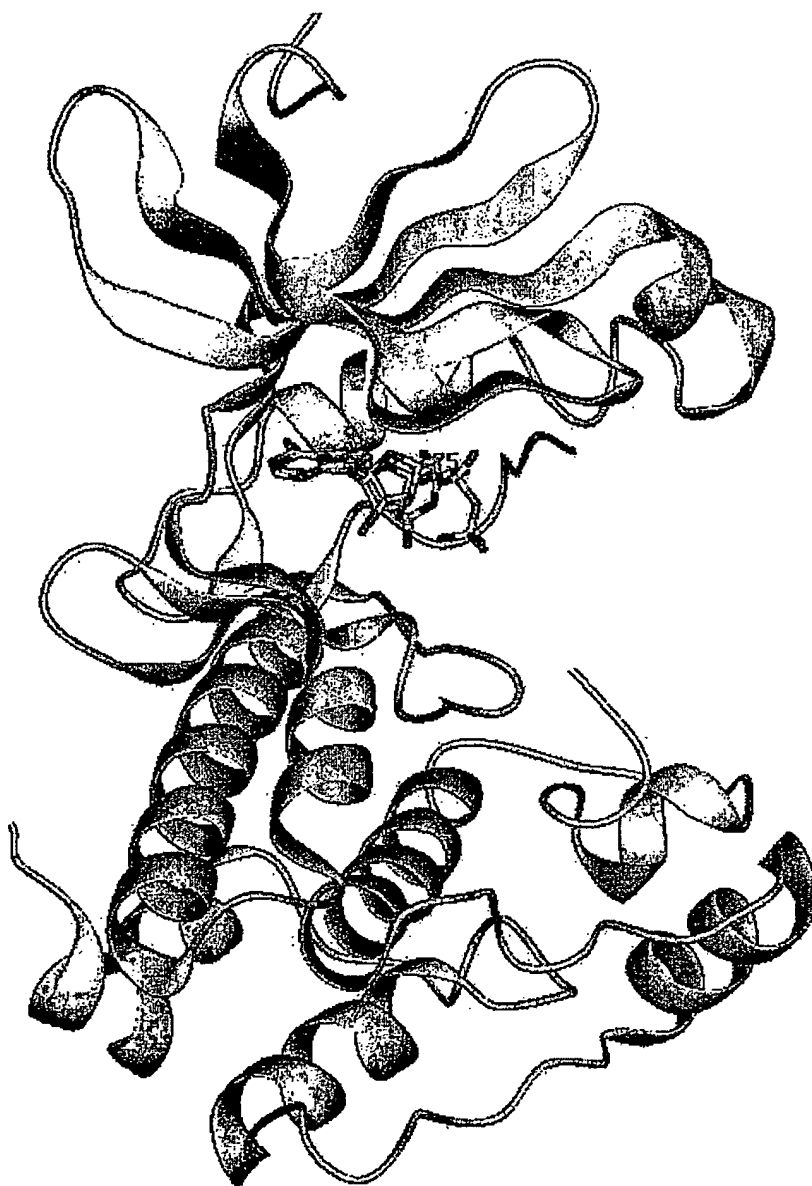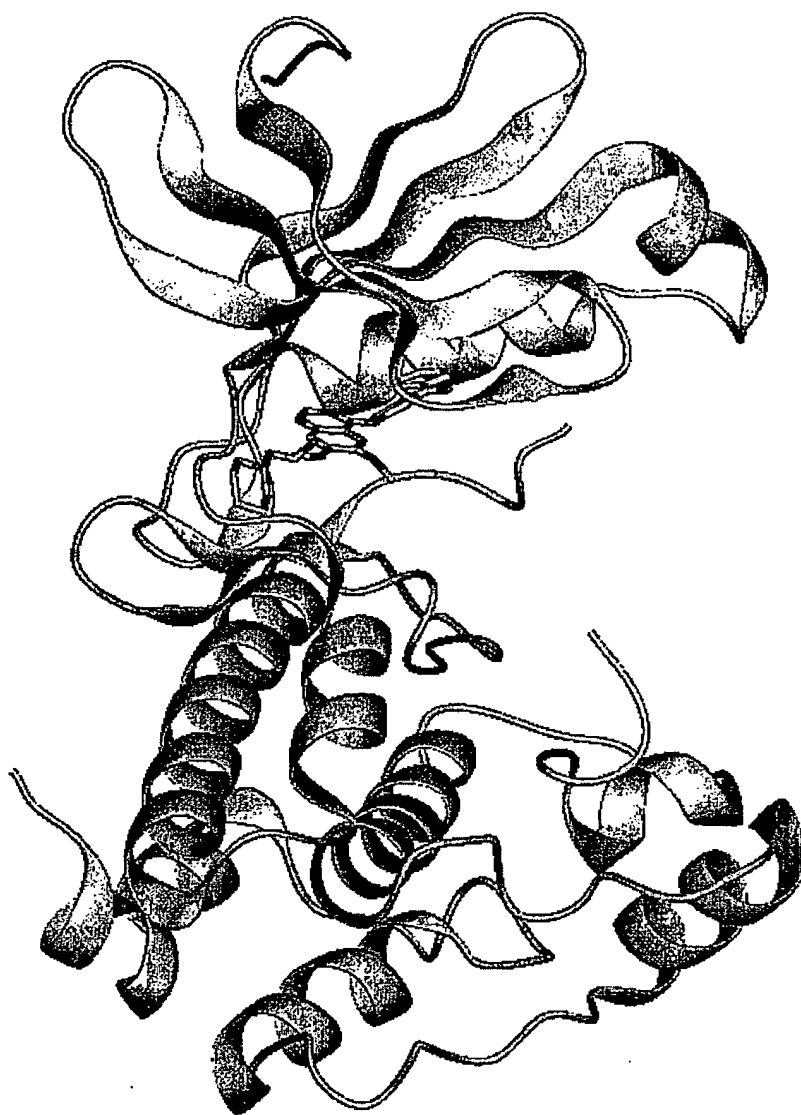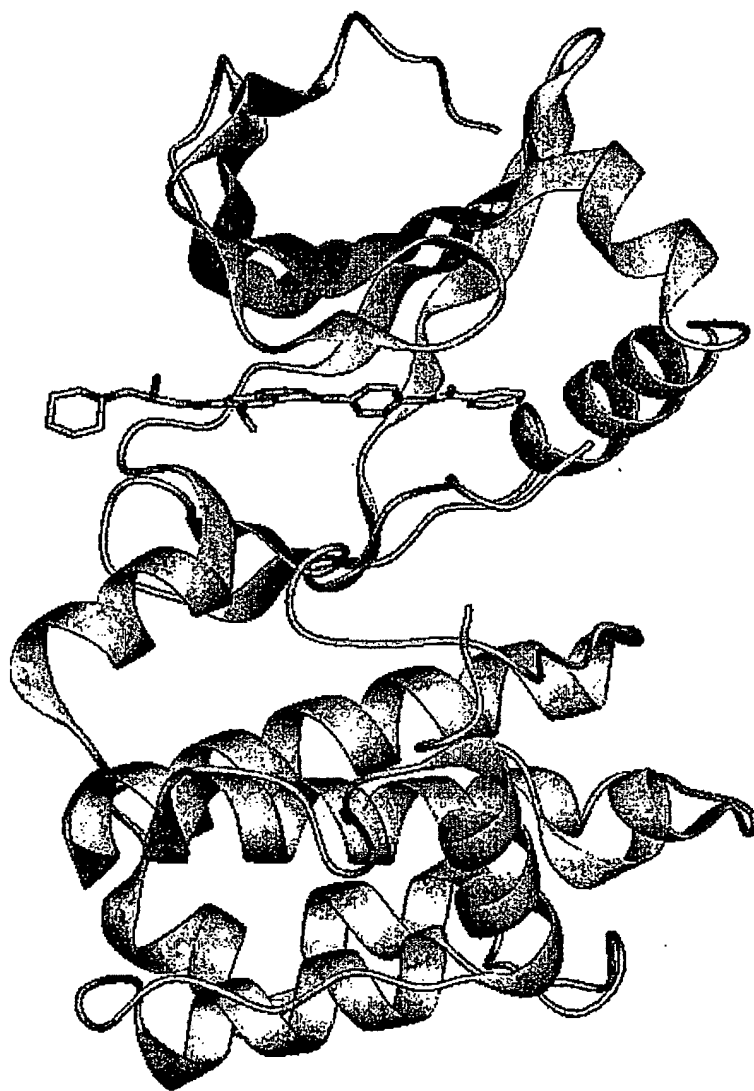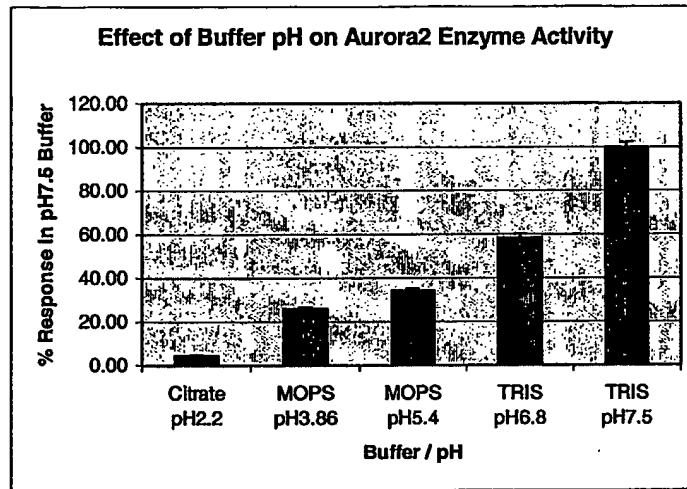
Figure 1

Figure 2a

Figure 2b

Figure 3